



# UNITED STATES PATENT AND TRADEMARK OFFICE

UNITED STATES DEPARTMENT OF COMMERCE  
United States Patent and Trademark Office  
Address: COMMISSIONER FOR PATENTS  
P.O. Box 1450  
Alexandria, Virginia 22313-1450  
www.uspto.gov

APPLICATION NO.	FILING DATE	FIRST NAMED INVENTOR	ATTORNEY DOCKET NO.	CONFIRMATION NO.
09/671,120	09/28/2000	Eiichi Takahashi	21.1980/CJG	8624

21171 7590 03/24/2004  
STAAS & HALSEY LLP  
SUITE 700  
1201 NEW YORK AVENUE, N.W.  
WASHINGTON, DC 20005

EXAMINER

SHARON, AYAL I

ART UNIT PAPER NUMBER

2123

DATE MAILED: 03/24/2004

2

Please find below and/or attached an Office communication concerning this application or proceeding.

h

## Office Action Summary

**Application No.**

09/671,120

**Applicant(s)**

TAKAHASHI ET AL.

**Examiner**

Ayal I Sharon

**Art Unit**

2123

-- The MAILING DATE of this communication appears on the cover sheet with the correspondence address --

### Period for Reply

A SHORTENED STATUTORY PERIOD FOR REPLY IS SET TO EXPIRE 3 MONTH(S) FROM THE MAILING DATE OF THIS COMMUNICATION.

- Extensions of time may be available under the provisions of 37 CFR 1.136(a). In no event, however, may a reply be timely filed after SIX (6) MONTHS from the mailing date of this communication.
- If the period for reply specified above is less than thirty (30) days, a reply within the statutory minimum of thirty (30) days will be considered timely.
- If NO period for reply is specified above, the maximum statutory period will apply and will expire SIX (6) MONTHS from the mailing date of this communication.
- Failure to reply within the set or extended period for reply will, by statute, cause the application to become ABANDONED (35 U.S.C. § 133). Any reply received by the Office later than three months after the mailing date of this communication, even if timely filed, may reduce any earned patent term adjustment. See 37 CFR 1.704(b).

### Status

- 1) ☐ Responsive to communication(s) filed on 28 September 2000.
- 2a) ☐ This action is **FINAL**. 2b) ☒ This action is non-final.
- 3) ☐ Since this application is in condition for allowance except for formal matters, prosecution as to the merits is closed in accordance with the practice under *Ex parte Quayle*, 1935 C.D. 11, 453 O.G. 213.

### Disposition of Claims

- 4) ☒ Claim(s) 1-14 is/are pending in the application.
- 4a) Of the above claim(s) \_\_\_\_\_ is/are withdrawn from consideration.
- 5) ☐ Claim(s) \_\_\_\_\_ is/are allowed.
- 6) ☒ Claim(s) 1-14 is/are rejected.
- 7) ☐ Claim(s) \_\_\_\_\_ is/are objected to.
- 8) ☐ Claim(s) \_\_\_\_\_ are subject to restriction and/or election requirement.

### Application Papers

- 9) ☐ The specification is objected to by the Examiner.
- 10) ☐ The drawing(s) filed on \_\_\_\_\_ is/are: a) ☐ accepted or b) ☐ objected to by the Examiner.  
Applicant may not request that any objection to the drawing(s) be held in abeyance. See 37 CFR 1.85(a).  
Replacement drawing sheet(s) including the correction is required if the drawing(s) is objected to. See 37 CFR 1.121(d).
- 11) ☐ The oath or declaration is objected to by the Examiner. Note the attached Office Action or form PTO-152.

### Priority under 35 U.S.C. § 119

- 12) ☒ Acknowledgment is made of a claim for foreign priority under 35 U.S.C. § 119(a)-(d) or (f).
- a) ☒ All b) ☐ Some \* c) ☐ None of:
1. ☒ Certified copies of the priority documents have been received.
2. ☐ Certified copies of the priority documents have been received in Application No. \_\_\_\_\_.
3. ☐ Copies of the certified copies of the priority documents have been received in this National Stage application from the International Bureau (PCT Rule 17.2(a)).
- \* See the attached detailed Office action for a list of the certified copies not received.

### Attachment(s)

- 1) ☒ Notice of References Cited (PTO-892)
- 2) ☐ Notice of Draftsperson's Patent Drawing Review (PTO-948)
- 3) ☐ Information Disclosure Statement(s) (PTO-1449 or PTO/SB/08)  
Paper No(s)/Mail Date \_\_\_\_\_.
- 4) ☐ Interview Summary (PTO-413)  
Paper No(s)/Mail Date. \_\_\_\_\_.
- 5) ☐ Notice of Informal Patent Application (PTO-152)
- 6) ☐ Other: \_\_\_\_\_.

## DETAILED ACTION

### *Introduction*

1. Claims 1-14 of U.S. Application 09/671,120 filed on 09/28/2000 are presented for examination. The application claims foreign priority to Japanese application 11-279516, filed on 09/30/1999.

### *Claim Rejections - 35 USC § 112*

2. The following is a quotation of the second paragraph of 35 U.S.C. 112:

The specification shall conclude with one or more claims particularly pointing out and distinctly claiming the subject matter which the applicant regards as his invention.

3. Claims 8-11 are rejected under 35 U.S.C. 112, second paragraph, as being indefinite for failing to particularly point out and distinctly claim the subject matter which applicant regards as the invention.
4. Claim 8 recites the limitation "for which a result of a ratio for which  $\beta$  multiplied by the standard value is less than or equal to  $\gamma$ ." in the last line. There is insufficient antecedent basis for the  $\beta$  and  $\gamma$  parameters in the claim.
5. Claim 9 recites the limitation "for which a result of ratio for which  $\beta$  multiplied by the standard value is smallest." in the last line. There is insufficient antecedent basis for the  $\beta$  parameter in the claim.
6. Claim 10 recites the limitation (emphasis added): "wherein said service modeling module categories each session transmission as a connection request and

response, and a command transmission, a data transmission, a response, ...".

The relationship between "connection response" and "response" is not clear.

7. Claim 11 recites the limitation (emphasis added): "wherein the permissions of each of the servers are taken as weighted values of a service distribution or relative ratios of the permissions are taken as server distribution ratios." It is not clear whether they are weighted in all cases or not.

### ***Claim Rejections - 35 USC § 102***

8. The following is a quotation of the appropriate paragraphs of 35 U.S.C. 102 that form the basis for the rejections under this section made in this Office action:

A person shall be entitled to a patent unless –

(b) the invention was patented or described in a printed publication in this or a foreign country or in public use or on sale in this country, more than one year prior to the date of application for patent in the United States.

9. The prior art used for these rejections is as follows:
10. Pitkin et al., U.S. Patent 5,341,477. (Henceforth referred to as "Pitkin").
- 11. Claims 1-2, 6-7 and 12-14 are rejected under 35 U.S.C. 102(b) as being anticipated by Pitkin.**

12. In regards to Claim 1, Pitkin teaches the following limitations:

1. A service distribution device for distributing services among a plurality of servers on a network to balance the server loads, comprising:  
a packet capture device capturing packets transmitted through the network;  
(Pitkin, especially: "Server connection section", Fig.4, Item 75, and col.6, line 54 to col.7, line 7; "Client connection section", Fig.4, Item 70, and col.9, lines 21-37. Pitkin's broker's capability to receive client requests and to poll servers inherently means that it captures packets transmitted through the network.)

a server identifier recording information pertaining to the captured packets into a server log for each server;

Art Unit: 2123

(Pitkin, especially: "Connection entries" in the "Server Status Block", Fig.4, Items 922-926, and col.6, line 54 to col.7, line 7. The Connection Entries store the status of the servers.)

a service identifier recording information pertaining to the captured packets into a service log for each service;

(Pitkin, especially: "Service List", Fig.4, Item 71, and col.6, lines 36-47, and col.6, line 65 to col.7, line 7. A service list is generated for those services found in repository - Fig.4, Item 42)

a server modeling module setting up the simulation model for each server from the server records;

(Pitkin, especially: "II. Modeling", col.5, lines 1-67, and Fig.3 teaches that: "A modeling process is used to efficiently determine the allocation of resources when designing a network" and "Both the server parameters and required service characteristics are inputs to a modeling process ..." and "The modeling process produces a prediction of the measured performance of each server based on the performance desired for each service offered by that server.")

a service modeling module setting up the simulation model for each service from the service records;

(Pitkin, especially: "II. Modeling", col.5, lines 1-67, and Fig.3 teaches that: "A modeling process is used to efficiently determine the allocation of resources when designing a network" and "Both the server parameters and required service characteristics are inputs to a modeling process ..." and "The modeling process produces a prediction of the measured performance of each server based on the performance desired for each service offered by that server.")

a simulator reading in the server model and the service model and running each simulation; and

(Pitkin, especially: "II. Modeling", col.5, lines 1-67, and Fig.3 teaches that: "A modeling process is used to efficiently determine the allocation of resources when designing a network" and "Both the server parameters and required service characteristics are inputs to a modeling process ..." and "The modeling process produces a prediction of the measured performance of each server based on the performance desired for each service offered by that server.")

a server selection module selecting and specifying an optimum server to distribute services to based on a simulator result.

(Pitkin, especially: "II. Modeling", col.5, lines 1-67, and Fig.3 teaches that: "A modeling process is used to efficiently determine the allocation of resources when designing a network" and "Both the server parameters and required service characteristics are inputs to a modeling process ..." and "The modeling process produces a prediction of the measured performance of each server based on the performance desired for each service offered by that server.")

13. In regards to Claim 2, Pitkin teaches the following limitations:

2. The service distribution device of claim 1, further comprising a packet relay device obtaining packets using a packet capture module mounted on said packet relay device, which relays packets between a client and the servers.

(Pitkin, especially: "Service List", Fig.4, Item 71, and col.6, lines 36-47, and col.6, line 65 to col.7, line 7; and "Connection entries" in the "Server Status Block", Fig.4, Items 922-926, and col.6, line 54 to col.7, line 7. The two together act as a device that relays packets between clients and servers.)

14. In regards to Claim 6, Pitkin teaches the following:

6. The service distribution device of claim 1, wherein said server selection module determines a standard value using an output of a single simulation run for each service by said simulator, and determines that a high load state exists if a difference between, or the ratio of, the standard value and the output of the simulation of a plurality of sessions exceeds a predetermined threshold.

(Pitkin, especially: col.9, lines 12-20: "By using these scan weight values, the servers having the greatest capacity, i.e. 22 and 24, will be assigned to multiple clients before being removed from the preview window. Thus, the client requests are evenly distributed across all available servers 21-25 based on their capacity." And also co.9, lines 52-60: ")

15. In regards to Claim 7, Pitkin does not expressly teach the following:

7. The service distribution device of claim 6, wherein when said server selection module receives a server distribution query, said server selection module sets a server permission to be a starting frequency of the session that will cause a high load state for the service in question for each server, and specifies a server having the biggest difference between the session starting frequency and the permission as a server for distribution.

(Pitkin, especially: col.9, lines 12-20: "By using these scan weight values, the servers having the greatest capacity, i.e. 22 and 24, will be assigned to multiple clients before being removed from the preview window. Thus, the client requests are evenly distributed across all available servers 21-25 based on their capacity." And also co.9, lines 52-60: ")

16. In regards to Claim 12, Pitkin teaches the following limitations:

12. A service distribution device for distributing services among a plurality of servers to balance server loads, comprising:

a server modeling module generating a simulation model for each server and a service modeling module generating a simulation model for each service;

(Pitkin, especially: "II. Modeling", col.5, lines 1-67, and Fig.3 teaches that: "A modeling process is used to efficiently determine the allocation of resources when designing a network" and "Both the server parameters and required service characteristics are inputs to a modeling process ..." and "The modeling process produces a prediction of the measured performance of each server based on the performance desired for each service offered by that server.")

Art Unit: 2123

a simulator reading the server models and the service models and running a plurality of simulations; and

(Pitkin, especially: "II. Modeling", col.5, lines 1-67, and Fig.3 teaches that: "A modeling process is used to efficiently determine the allocation of resources when designing a network" and "Both the server parameters and required service characteristics are inputs to a modeling process ..." and "The modeling process produces a prediction of the measured performance of each server based on the performance desired for each service offered by that server.")

a server selection module determining which servers have low loads based on results of the simulations and selecting the servers with low loads to receive the services.

(Pitkin, especially: col.9, lines 12-20: "By using these scan weight values, the servers having the greatest capacity, i.e. 22 and 24, will be assigned to multiple clients before being removed from the preview window. Thus, the client requests are evenly distributed across all available servers 21-25 based on their capacity." And also co.9, lines 52-60: ")

17. In regards to Claim 13, Pitkin teaches the following limitations:

13. A method for distributing services among a plurality of servers to balance server loads, comprising:

generating a simulation model for each server and each service;

(Pitkin, especially: "II. Modeling", col.5, lines 1-67, and Fig.3 teaches that: "A modeling process is used to efficiently determine the allocation of resources when designing a network" and "Both the server parameters and required service characteristics are inputs to a modeling process ..." and "The modeling process produces a prediction of the measured performance of each server based on the performance desired for each service offered by that server.")

running a plurality of simulations using the server and service models; and

(Pitkin, especially: "II. Modeling", col.5, lines 1-67, and Fig.3 teaches that: "A modeling process is used to efficiently determine the allocation of resources when designing a network" and "Both the server parameters and required service characteristics are inputs to a modeling process ..." and "The modeling process produces a prediction of the measured performance of each server based on the performance desired for each service offered by that server.")

determining which servers have low loads based on results of the simulations

(Pitkin, especially: "II. Modeling", col.5, lines 1-67, and Fig.3 teaches that: "A modeling process is used to efficiently determine the allocation of resources when designing a network" and "Both the server parameters and required service characteristics are inputs to a modeling process ..." and "The modeling process produces a prediction of the measured performance of each server based on the performance desired for each service offered by that server.")

and selecting the servers with low loads to receive the services.

(Pitkin, especially: col.9, lines 12-20: "By using these scan weight values, the servers having the greatest capacity, i.e. 22 and 24, will be assigned to multiple clients before being removed from the preview window. Thus, the client requests are evenly distributed across all available servers 21-25 based on their capacity." And also co.9, lines 52-60: ")

18. In regards to Claim 14, Pitkin teaches the following limitations:

14. A computer-readable storage controlling a computer and comprising a process of:  
generating a simulation model for each server and each service;

(Pitkin, especially: "II. Modeling", col.5, lines 1-67, and Fig.3 teaches that: "A modeling process is used to efficiently determine the allocation of resources when designing a network" and "Both the server parameters and required service characteristics are inputs to a modeling process ..." and "The modeling process produces a prediction of the measured performance of each server based on the performance desired for each service offered by that server.")

running a plurality of simulations using the server and service models; and

(Pitkin, especially: "II. Modeling", col.5, lines 1-67, and Fig.3 teaches that: "A modeling process is used to efficiently determine the allocation of resources when designing a network" and "Both the server parameters and required service characteristics are inputs to a modeling process ..." and "The modeling process produces a prediction of the measured performance of each server based on the performance desired for each service offered by that server.")

determining which servers have low loads based on results of the simulations and

(Pitkin, especially: "II. Modeling", col.5, lines 1-67, and Fig.3 teaches that: "A modeling process is used to efficiently determine the allocation of resources when designing a network" and "Both the server parameters and required service characteristics are inputs to a modeling process ..." and "The modeling process produces a prediction of the measured performance of each server based on the performance desired for each service offered by that server.")

selecting the servers with low loads to receive the services.

(Pitkin, especially: col.9, lines 12-20: "By using these scan weight values, the servers having the greatest capacity, i.e. 22 and 24, will be assigned to multiple clients before being removed from the preview window. Thus, the client requests are evenly distributed across all available servers 21-25 based on their capacity." And also co.9, lines 52-60: ")



Art Unit: 2123

19. The following is a quotation of 35 U.S.C. 103(a) which forms the basis for all obviousness rejections set forth in this Office action:

(a) A patent may not be obtained though the invention is not identically disclosed or described as set forth in section 102 of this title, if the differences between the subject matter sought to be patented and the prior art are such that the subject matter as a whole would have been obvious at the time the invention was made to a person having ordinary skill in the art to which said subject matter pertains. Patentability shall not be negated by the manner in which the invention was made.

20. The prior art used for these rejections is as follows:

21. Pitkin et al., U.S. Patent 5,341,477. (Henceforth referred to as "**Pitkin**").

22. Abbott et al., U.S. Patent 6,314,463. (Henceforth referred to as "**Abbott**").

23. Kleinrock, L. "On the Modeling and Analysis of Computer Networks." Proc. of the IEEE. Aug. 1993. pp.1179-1191. (Henceforth referred to as "**Kleinrock**").

24. Microsoft Press Computer User's Dictionary. © 1998. p.344. (Henceforth referred to as "**Microsoft**").

25. Jain, R. The Art of Computer Systems Performance Analysis. © 1991. pp.624-626. Specifically, the section titled "Symbols Frequently Used in Queueing Analysis." (Henceforth referred to as "**Jain**").

26. The claim rejections are hereby summarized for Applicant's convenience. The detailed rejections follow.

27. Claim 3 is rejected under 35 U.S.C. 103(a) as being unpatentable over Pitkin in view of Microsoft and further in view of Abbott.

28. Claim 4 is rejected under 35 U.S.C. 103(a) as being unpatentable over Pitkin in view of Jain and further in view of Kleinrock.

29. Claims 5 is rejected under 35 U.S.C. 103(a) as being unpatentable over Pitkin in view of Kleinrock.

30. In regards to Claim 3, Pitkin does not expressly teach the following limitations:

3. The service distribution device of claim 1,

wherein said server modeling module constructs a server model having a queue corresponding to a transmission process using the server log and a server transmission throughput, a server processing time, and a unit processing time as parameters,

wherein the server transmission throughput is calculated from a total size  $L$  of an arbitrary, continuous string of the continuously transmitted packets using the formula  $L / (t_e - t_s)$  where  $t_e$  is an ending packet capture time and  $t_s$  is a starting packet capture time, and

wherein the server processing time is calculated using the formula  $(t_s - t_c) - (l_s + l_c) / B$ , wherein  $t_s$  and  $l_s$  are the capture time and size of a server response packet, respectively,  $t_c$  and  $l_c$  are the capture time and size of a corresponding client response packet, respectively, and  $B$  is a network speed.

Microsoft, on the other hand, teaches that "throughput" is a "... measure of ... the data processing rate in a computer system." By definition, this measurement of (quantity of data / period of time).

It would have been obvious to one of ordinary skill in the art at the time the invention was made to modify the teachings of Pitkin with those of Microsoft, because the Microsoft dictionary teachings were well known in the art at the time the invention was made.

Abbott teaches (see Fig.5) that order to calculate the response time, it is not sufficient to subtract the "request sent" time stamp (at the client) from the "end of processing" time stamp (at the server). The queue delay must also be included in the calculations. Inherently, the queue delay also applies to when the server sends the processing results to the client.

This corresponds to Applicant's claimed formula. The formula component " $(t_s - t_c)$ " corresponds to subtracting the "request sent" time stamp (at the client)

from the "end of processing" time stamp (at the server). The queue delay (network delay) corresponds to the formula component  $(I_s + I_c) / B$ .

It would have been obvious to one of ordinary skill in the art to modify the teachings of Pitkin with those of Abbott, because Abbott's tools help to "coordinate the operation of multiple web servers." (Abbott, Abstract).

31. In regards to Claim 4, Pitkin does not expressly teach the following limitations:

4. The server distribution device of claim 1, wherein said service modeling module calculates the following parameters from the service log by constructing a service model for each service:

a ratio of the number of sessions for each service to the number of sessions for all services,

a session starting frequency or time interval,

a number of transmissions between the client and server per session,

a client response size, packet size, and packet count per transmission,

a server response size, packet size, and packet count per transmission, and

a time from the server response until the client response.

Jain, on the other hand, teaches the following parameters which correspond to those claimed by the Applicants:

- The ratio of the parameters  $D_i$  and  $D$ , (Total service demand on server 'i', and Total service demand on all servers), as taught by Jain, produces a ratio that corresponds to Applicants' claimed limitation:

a ratio of the number of sessions for each service to the number of sessions for all services,

- The parameter  $\tau$ , (Inter-arrival time), as taught by Jain, corresponds to Applicant's claimed limitation:

a session starting frequency or time interval,

- The parameter  $V_i$ , (Number of visits to service center  $i$ ), as taught by Jain, corresponds to Applicant's claimed limitation:

a number of transmissions between the client and server per session,

- The parameter  $I$ , (Idle time duration for a server), as taught by Jain, corresponds to Applicant's claimed limitation:

a time from the server response until the client response.

It would have been obvious to one of ordinary skill in the art to modify the teachings of Pitkin with those of Jain, because Jain's symbols are "Frequently Used in Queueing Analysis".

Kleinrock teaches the following parameters which correspond to those claimed by the Applicants:

a client response size, packet size, and packet count per transmission,

a server response size, packet size, and packet count per transmission, and

Kleinrock teaches (p.1180, col.2, para.3) that "The modification is to assume that message lengths are all the same (rather than the exponential assumption above), and that the topology is a tandem network.

It would have obvious to one of ordinary skill in the art at the time the invention was made to modify the teachings of Pitkin with those of Kleinrock, Because Pitkin expressly teaches (col.5, lines 27-32) that "Both the server parameters and required service characteristics are inputs to modeling process such as is described in [two other Kleinrock references related to modeling queuing computer networks]."

32. In regards to Claim 5, Pitkin does not expressly teach the following:

5. The service distribution device of claim 1, wherein said simulator performs a simulation using the server model and the service model and generates a mean value or a median value of a session time for the specific service.

Kleinrock, on the other hand, does expressly teach:

a) generating the mean delay time of a system "One of the first general results was an exact expression for the mean delay experienced by a message as it passed through a network ..." (See p.1179, col.2, paragraph 4).

b) generating the mean response time of a system "In addition, we let  $T_i$  be the mean response time of this little queueing system." (See p.1180, col.2, Eq.4)

It would have obvious to one of ordinary skill in the art at the time the invention was made to modify the teachings of Pitkin with those of Kleinrock, Because Pitkin expressly teaches (col.5, lines 27-32) that "Both the server parameters and required service characteristics are inputs to modeling process such as is described in [two other Kleinrock references related to modeling queuing computer networks]."

### ***Correspondence Information***

Any inquiry concerning this communication or earlier communications from the examiner should be directed to Ayal I. Sharon whose telephone number is (703) 306-0297. The examiner can normally be reached on Monday through Thursday, and the first Friday of a biweek, 8:30 am – 5:30 pm.

Art Unit: 2123

If attempts to reach the examiner by telephone are unsuccessful, the examiner's supervisor, Kevin Teska can be reached on (703) 305-9704. Any response to this office action should be mailed to:

Director of Patents and Trademarks  
Washington, DC 20231

Hand-delivered responses should be brought to the following office:

4<sup>th</sup> floor receptionist's office  
Crystal Park 2  
2121 Crystal Drive  
Arlington, VA

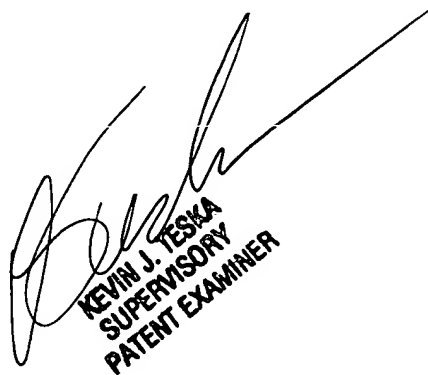
The fax phone number is: (703) 872-9306

Any inquiry of a general nature or relating to the status of this application or proceeding should be directed to the receptionist, whose telephone number is: (703) 305-3900.

Ayal I. Sharon

Art Unit 2123

March 12, 2004



KEVIN J. TESKA  
SUPERVISORY  
PATENT EXAMINER